METHOD AND ARRANGEMENT RELATING TO X-RAY IMAGING

Field of invention

5   The present invention relates to the detection of specific characteristics in an X-ray image, and more especially malignant tumors in digitally produced mammograms, and in particular to a method of finding stellate lesions based on phase information obtained from for instance quadrature filters.

10   *Background of invention*

Breast cancer is a serious health threat and effects many women each year. At the present, there is no existing means for preventing breast cancer; however methods have been developed for screening women for early detection of cancer.

15   Mammography using x-rays is currently the most used method and is used for screening large populations of people. It is of importance to diagnose patients at as an early stage as possible, which means that the malignant lesions are small and hard to detect.

20   The large quantity of people to screen means that a large amount of images has to be screened and a physician or radiologist may be required to examine several hundreds of mammograms per day. This increases the risk of a missed diagnosis due to human error especially as the lesions may be small and hard to detect.

25   Accordingly, Computer Aided Diagnosis (CAD) systems for screening of medical digital images have been developed for assisting in the detection of abnormal lesions, for instance spiculations. Malignant lesions can often be revealed by looking for spiculations, i.e. stellar-shaped lesions. These may be visible in mammograms and come in many different sizes. The presence of stellate-like spicules radiating

30   from a center mass is a highly suspicious indicator of malignancy. Many methods and systems have therefore been developed for the detection of such features in x-ray images.

Karssemeijer *et al* (N. Karssemeijer *et al*, "Detection of Stellate Distortions in

35   Mammograms", IEEE transactions on Medical Imaging, Vol 15, No 5, pp 611-619, 1996) suggested a statistical method based on a map of pixel orientations. Another method is based on first identifying individual spicules and then via a Hough

transform, accumulates evidence that they point in a certain direction. This method is used for instance by Kobatake *et al* (H. Kobatake *et al*, "Detection of Spicules on Mammogram Based on Skeleton Analysis", IEEE Transactions on Medical Imaging, Vol. 15, No 3, pp 235-245) and Ng *et al* (S. L. Ng *et al*, "Automated detection and classification of breast tumors", Computers and Biological Res., Vol. 25, pp 218-237, 1992.

A third method is based on histogram analysis of gradient angles as proposed in Kegelmeyer (W. P. Kegelmeijer Jr., "Computer Detection of Stellate Lesions in Mammograms", Proc. SPIE Conf. Biomedical Image Processing and Three-Dimensional Microscopy, Vol 1660, 1992). The basic idea is that if the standard deviation of gradient angles in a certain local neighborhood or area is high, then it is an indication that the gradients point in all-different directions. This would indicate a stellate pattern. This is also outlined in US 5,633,958, wherein a method and apparatus for detecting a desired behavior in digital image data is presented. In this system stellate lesions are detected in digitized mammography image data using an ALOE (analysis of local oriented edges) approach is implemented to calculate features. The primary disadvantage of using the ALOE algorithm is that many unwanted background objects can produce signals false signals indicative of malignant lesions. Also because every direction may not be present in the histogram of gradient angles, the standard deviation of the histogram may still be quite large resulting in a larger ALOE signal and spiculations may thus be missed. Thus the ALOE algorithm produces false positives and also results in missed spiculations.

A common problem when detecting spiculated lesions is that they range in size from a few millimeters up to several centimeters. This may be problematic for some lesion detection methods. One way of addressing this problem is to use the detection system on several different scales. Karssemejer *et al* uses this kind of approach to overcome this problem.

Another solution for finding lesions in images is based on an artificial neural network that compares found features in an unknown image with features found in images with known diagnoses and this solution is presented in US patent application number 2001/0043729. Since this is based on the availability of images of known diagnoses it will only find similar looking lesions.

Yet another solution is presented in US 6,263,092, wherein a method and apparatus for fast detection of spiculated lesions using line and direction information found in the image and accumulating regions of possible intersections to produce a cumulative array. Information derived from the cumulative array is

5   used for identifying spiculations in the digital mammogram image. One problem with this method is that both stellar and circle shaped features will result in the similar histograms and thus the method will produce false positive signals increasing the burden on the radiologist/physician that manually interpret and examine the images before diagnosing.

10

Summary of invention

The present invention proposes a novel method and apparatus for detecting interesting characteristics in an x-ray image, and more especially malignant lesions

15   or suspicious features in digital medical images and in particular proposes a new method for finding the Region of Interest (ROI) in a CAD (Computer Aided Diagnosis) system that has many optimization possibilities and yet is fast and accurate and still overcomes some of the above mentioned problems.

20   For these reasons, a method for detection of stellate lesions in a digitalized mammogram is provided. The method comprises the steps of: obtaining an image data corresponding to the mammogram; obtaining an image mask; substantially uniformly sampling the digital image inside the mask and producing sample points; calculating for each sample point a characteristic; selecting a number of sampling

25   points most likely to correspond to a spiculated lesion; applying a segmentation procedure to the original digital image at the selected sampling points; extracting new characteristics from each segmented area and obtaining a feature vector; classifying each feature vector as suspicious or non-suspicious using a classification machine; and examining the suspicious areas. The characteristics comprise one or

30   several of: contrast, two measures of spiculatedness, and two measures of edge orientations. The contrast is derived as a ratio between intensity inside a circle with a radius r1 and a washer shaped background area with inner radius r1 and an outer radius r2. The two measures of spiculatedness are derived from a histogram of angle differences obtained using a filtration method that yields phase information

35   together with orientation estimates. The two measures of edge orientations are derived from a histogram of angle differences obtained using a filtration method that yields phase information together with orientation estimates.

Extracting can be done using a support vector machine or an artificial neural network. The classification of each feature vector can be done using a classification machine. Preferably, the entire image is sampled. Each node in the applied 5 sampling grid is evaluated in terms of contrast and spiculation.

The invention also relates to a method of detecting a Region of Interest in a digitalized X-ray image, comprising the steps of: extracting phase information from the image, using the phase information for differentiating between different lines 10 and edges, and skewing the lines towards a centre. The first step comprises extracting an orientation estimate. The second step comprises additional information on a magnitude from a filter answer.

The invention also relates to an arrangement for detecting a Region of Interest in a 15 digitalized X-ray image. The arrangement comprises: a processing unit, a module for obtaining image masks, a sampling module, a calculating module, filtration module, a classification module and a support vector machine and/or artificial neural network module. The filtration module is a set of quadrature-filter. The invention also relates to n x-ray apparatus comprising an above-mentioned 20 arrangement.

The invention also relates to a computer unit comprising a processing unit, a memory unit, storage unit, the computer unit being operatively arranged with an instruction set to acquire a digitalized x-ray image. The instruction set has 25 procedures for: detecting a Region of Interest in a digitalized X-ray image, extracting phase information from the image, obtaining image masks, sampling, calculating, filtration, a classification and supporting vector and/or artificial neural network.

30 The invention may be realized as a computer program for detection of stellate lesions in a digitalized mammogram. The program comprises: an instruction set for obtaining an image data corresponding to the mammogram; an instruction set for obtaining an image mask; an instruction set for substantially uniformly sampling the digital image inside the mask and producing sample points; a calculation 35 procedure for each sample point a characteristic; an instruction set for selecting a number of sampling points most likely to correspond to a spiculated lesion; an instruction set for applying a segmentation procedure to the original digital image

at the selected sampling points; an instruction set for extracting new characteristics from each segmented area and obtaining a feature vector; and classifying procedure for classifying each feature vector as suspicious or non-suspicious using a classification machine.

5

Brief description of drawings

.The present invention will become more fully understood from the detailed description given below together with the accompanying drawings, which are given
10   for illustrative purposes only and should not be considered limiting the present invention and wherein:

Fig. 1   illustrates an X-ray apparatus employing an arrangement according to the present invention.

15

Fig. 2   A and B illustrates an image area with a malignant lesion and a corresponding line image of the same area respectively.

Fig. 3   A and B illustrates an image area with a malignant lesion and a
20   corresponding edge image of the same area respectively.

Fig. 4   A and B shows histograms of angle difference distributions from line and edge analysis respectively.

25   Fig. 5   A and B shows original mammogram and SVM output from a ROI extraction step respectively. A stellate lesion is marked in both images with an arrow (A and B).

Fig. 6   A shows a local neighborhood of a malignant stellate lesion and B shows
30   the output from a level set segmentation algorithm for the same local neighborhood.

Fig. 7   A and B shows an original grid output and SVM output after segmentation respectively.

35

Fig. 8   shows block diagram of an arrangement implementing the stellate lesion detection method of the present invention.

Fig. 9   is flow diagram illustrating the main steps of the invention.

Fig. 10   is distribution of the angle differences corresponding to the pixels.

5   *Detailed description of invention*

The present invention proposes a novel method for detecting Region of Interests
with special characteristics generally and particular stellate lesions in digitized x-
rays images, especially mammogram images, in the scope of computer-aided
diagnosis (CAD). The method/system is used as an aid to radiologists or physicians

10   in the characterization and classification of mass lesions in mammography. Studies
have shown that such a system can aid in increasing the diagnostic accuracy and
increase the examination rate. According to the most general implementation, the
invention comprises detecting a Region of Interest in a digitalized X-ray image by:
extracting phase information from the image, using the phase information for

15   differentiating between different lines and edges, and skewing the lines towards a
centre. The extraction step comprises extracting a orientation estimate. The phase
information comprises additional information on a magnitude from a filter answer.

An exemplary X-ray apparatus is illustrated in a schematic way in Fig. 1. The

20   apparatus 100 comprises an x-ray source 110, a collimator 120 and a detector
assembly 130 arranged in a housing 140 and supported by a supporting structure
101. The housing further comprises an upper plate 141 housing the collimator 120
and a lower plate 142 housing the detector assembly 130. An object to be
examined, e.g. a breast, is positioned between the upper and lower plates and

25   compressed before exposure to the X-rays. In this case a computer 150 is
connected to the X-ray apparatus for processing the information received from the
detector assembly, e.g. execute CAD.

A CAD method according to the present invention includes several steps with

30   different purposes and these will be presented in conjunction with Fig. 9 in an order
as they appear in the process.

The first step involves obtaining a digital image 901 from a mammography
measurement, e.g. the aforementioned apparatus 100. The image may be obtained

35   directly from the X-ray apparatus, scanning a film obtained during a mammography
measurement (film based mammography apparatus), or collecting an image from a
database of stored images located either locally at a mammography facility or

externally at some central database. For instance for test, training, and evaluation purposes, images may be obtained from the Digital Database for Screening Mammography at the University of South Florida, etc.

5   In some cases the images need some image pre-processing, for instance noise reduction or thickness equalization, before starting the actual detection algorithm.

Preferably, the image is subjected 902 to a mask according to standard tools in the field.

10

The mammogram is subjected to a grid pattern in order to uniformly sample 903 the image inside the mask. This is done by applying the grid with a distance d between nodes in x and y directions.

15   For each sampling point obtained above, several features are calculated 904:

   i)   The contrast of the image is calculated by calculating the ratio between the average intensity inside a circle with radius r1 and a washer shaped background area with inner radius r1 and outer radius r2,

   ii)   Two measures of so called spiculatedness is derived from a histogram of
20   angle differences which will be discussed in more detail below, and

   iii) Two measures of edge orientations are derived from the histogram of angle differences.

A support vector machine or any other learning machine such as an artificial neural
25   network may be used to select 905 a number of sampling points that are most likely to correspond to malignant tissue, in particular spiculated lesions. A segmentation algorithm is applied 906 to the original mammogram at coordinates corresponding to the current sampling point as is illustrated in Fig. 6 in order to prevent sampling points close to each other from being extracted and to use the
30   segmented area to extract refined features.

New features are extracted from each segmented area, including, but not limited to, contrast between the segmented Region of Interest (ROI) and its immediate background, spiculation and edge measures calculated using the same method as
35   above, texture features are calculated according standard tools in the technical field, shape features are also calculated using standard tools, and intensity based features are calculated using standard tools of the trade.

Each feature vector is passed on to a classifying machine to be classified into either suspicious or non-suspicious features. A user-defined threshold may be implemented in order to determine the trade off between false positive findings and

5   false negative findings.

Suspicious areas are marked for later examination by a radiologist or physician.

In the following, above described steps are detailed.

10

In order to find regions of interest (ROIs) different methods for finding seed points exist. Most methods are intensity based using the fact that many tumors have a well-defined central body, whereas other methods search for spiculation features and try to determine from where the spicules emanate from. The present invention

15   uses a combination of these two methods and adds another method to capture the edge orientation. The entire image is sampled in order to minimize the risk of missing any areas of interest and each node in the applied sampling grid is evaluated in terms of contrast and spiculation.

20   As mentioned before, the features vary in size and therefore this evaluation is done on three different scales.

The contrast measured at node i, j is defined as the contrast between a circular area with radius r1 centered at i, j and a washer shaped area with inner radius r1

25   and outer radius r2. r1 and r2 can be any size but may for instance be r and $2r$.

The spiculation and edge measures are based on orientation estimates extracted from a filtration method that can extract phase information together with orientation estimates. One such filtration method may be for instance by using a

30   quadrature filter set, e.g. four filters.

An example employing a quadrature filter is disclosed in the following:

Quadrature filters and a method to construct orientation tensors from the

35   quadrature filter are described in G. H. Granlund, H. Knutsson, "Signal Processing for Computer Vision", Kluwer Academic Publishers, Dordrecht, 1995. The directing vector of quadrature filter i is denoted $\hat{n}_i$ with $\varphi_i = \arg(\hat{n}_i)$. The quadrature filter is

9

complex and hence the output q_i from convolution of the filter and the image signal will be complex. Let q_i denote the magnitude and $\boldsymbol{q_i}$ and similar for the phase angle $\theta_i = \arg(\boldsymbol{q_i})$.

5   The local orientation in an image is the direction in which the signal exhibits maximal variation. With $\theta_i = (i-1)* \pi/4$, the 2D orientation vector may be expressed conveniently as

$$Z = (q_1 - q_3,\; q_2 - q_4).$$

10

Thus, if $\boldsymbol{v}$ is a vector oriented along the axis of maximal signal orientation, the following relationship hold between the arguments of $\mathbf{z}$ and $\mathbf{v}$: $\arg(\mathbf{z}) = 2* \arg(\mathbf{v})$.

The phase angle introduced above reflects the relationship between the evenness

15   and oddness of the signal. In the spatial domain, a quadrature filter may be written as a sum of a real line detector and a real edge detector:

$$f(x) = f_{line}(x) - i\, f_{edge}(x).$$

20   f_{line} is an even function and f_{edge} is an odd function and this can be used to distinguish between lines and edges. Extending the phase concept to two dimensions is not trivial, but will give the necessary means to distinguish different features from each other, namely edges, bright lines, and dark lines. The reason for the difficulties is that the phase can not be defined independently of directions, and

25   as the directing vectors of the quadrature filters point in different directions, and thus yield opposite signs for similar events, care must be taken in the summation. A method for weighting the filter output is the following: let $\Re(q_i)$ and $\Im(q_i)$ denote the real and imaginary parts of the filter output from the quadrature filter in direction $\hat{n}_i$. The weighted filter output is then given by

30

$$\Re(q) = \sum_{i=1}^{4} \Re(q_i)$$

$$\Im(q) = \sum_{i=1}^{4} sign(\cos(\varphi_i - \varphi))\Im(q_i)$$

The interpretation of the cosine factor is that when the local orientation in the image and the directing vector of the filter differ by more than $\pi/2$ the filter output must be conjugated to account for the anti-symmetric imaginary part. The

5   total phase θ is now given as $\theta = \arg(q) = \arg(\Re(q) + i\Im(q))$. Phase angles close to zero correspond to bright lines, phase angles close to $\pm\pi$ correspond to dark lines and phase angles close to $\pm\ \pi/2$ correspond to edges.

By thresholding the filter outputs on certainty and phase, a line image is produced.
10 This may be used to separate bright lines and thus candidates for spicules, from the surrounding tissue. Such a test is shown in Fig. 2, where the real image is shown on the left 1A and the calculated image is shown to the right 1B using a particular phase angle threshold.

15 Using another phase angle threshold an edge image is produced as may be seen in Fig. 3, wherein 3A is the real image and 3B is the calculated image.

There is a clear difference in these two images 2B and 3B. The question now comes up on how to quantify this difference. This is achieved by constructing a measure of
20 spiculatedness in a local area or neighborhood. The direction of maximal signal variation in a pixel on a detected bright line is *v(x)* and let $\varphi = \arg(v(x))$. Then we get the following expression for the double angle representation of local orientation:

$$z(x) = c(x)e^{i2\varphi} = q_1 - q_3 + i(q_2 - q_4).$$

25

Let $\hat{r}$ denote a normalized vector pointing from a coordinate $\mathbf{x}_0$ in the image to another pixel **x**. Since the vector $\hat{r}$ is normalized it may be expressed as $(\cos\varphi_r(x), \sin\varphi_r(x))$. Let us now define

30 $\hat{r}_{double}(x) = (\cos(2\varphi_r), \sin(2\varphi_r))$.

If **x** is located on a line radiating away from the center coordinate, the angles between $\hat{r}_{double}$ and **z(x)** will be π. On the other hand, if **x** is located on a line perpendicular to $\hat{r}$, the angle will be zero. To see that, consider Fig. 10 where ψ

denotes the angle between $\hat{r}(x)$ and v(x). From the figure it is obvious that arg(v) = φ_r + ψ ± π/2. This means that

$$\text{arg}(z) = 2\varphi_r + 2\psi \pm \pi = 2\varphi_r + 2\psi + \pi \quad (\text{modulo } 2\pi)$$

Since arg($\hat{r}_{double}$) = $2\varphi_r$ the absolute value of the difference between the angles modulo 2π is

$$|\varphi| = \text{arg}(z) - \text{arg}(\hat{r}_{double}) = 2\psi \pm \pi \quad (\text{modulo } 2\pi).$$

Now, with ψ close to zero, as it would be if the line is part of a stellate pattern, the angle difference will be close to π, as proposed above. On the other hand, if the line is perpendicular to r the angle difference φ will be close to zero.

Thus, if the distribution of the angle differences corresponding to the pixels identified in the line image in a local neighborhood is skewed toward π as may be seen in Fig. 4A, this is an indication that many lines are radiating away from the center. If the pixel orientations of the edge image are skewed towards the left in the Fig. 4B, this is an indication that the prominent edges are perpendicular to lines radiating from the center.

The next step in the process is to apply the data to a ROI extractor. Five features are used in the ROI extractor: contrast as discussed above, two fraction of points in the line image in the washer shaped neighborhood that have particular angle deviations, and two features that are similar measures for the edge image.

A support vector machine (SVM) or similar learning machine such as an artificial neural network is used to distinguish between areas that could be potentially malignant and those that could not. This learning machine has been trained using known data prior to using it on unknown data.

Image features (for example the five features mentioned above) are extracted in a number of locations in the image and since the size of possible lesions is unknown three different radii on the washer shaped area are evaluated. The radius where the corresponding features give the highest SVM response is taken as the size of ROI. A typical intermediate result of the ROI is illustrated in Fig. 5, wherein A shows a normal image and B an SVM output from the ROI extraction step.

It should be noted that Fig. 5B do not represent the final classifying decision of the CAD system, but rather the first step of localizing the ROIs that should be further processed. The coordinates with the highest response are then extracted and passed on to the segmentation step.

5

The coordinates with the highest intensity maxima are extracted as seen in Fig. 5B and a boundary refinement algorithm is initiated around this neighborhood for segmentation. There is several available boundary refinement algorithms may be used in this step. One illustration of the output of such an algorithm may be seen in

10   Fig. 6B using a level set segmentation boundary refinement algorithm, Fig. 6A displays the original digital medical image.

Once the ROI has been segmented from the background, its immediate background is determined as all pixels within a distance d from the ROI, where d is chosen such

15   that the area of background roughly corresponds to the area of the ROI and thus an extended ROI has been constructed. Then the extended ROI is removed from the ROI extractor grid output as shown in Fig. 7. Fig. 7A is the SVM output image and 7B represents a segmented SVM image. This process is repeated until a number of regions of interest are passed on to the next steps in the process: feature

20   extraction and classification.

Using the segmented results, the five features are recalculated using the segmented ROI and its immediate surrounding instead of the washer shaped neighborhoods used in the ROI extraction step. Some additional features are added

25   to aid in the classification. The standard deviation of the interior of the ROI normalized with the square root of the intensity yields a texture measure capturing the homogeneity of the area. An equivalent feature is extracted for the immediate background. The compactness of the segmented ROI is also extracted and these features are then passed on to a classifying machine. The same learning machine

30   implementation as mentioned above is trained with the features from these refined areas.

The final step involves marking the image at found suspicious areas and points for final examination of a radiologist or physician.

35

The method described above may be implemented in a dedicated external device or apparatus, or incorporated in a mammogram system.

It may also be implemented on a computer medium as a stand-alone system implemental in any computational device with sufficient computing power. Thus, the entire method or parts of the same can be provided as instruction set

5   (computer program).

An exemplary arrangement 800 for processing the image according to the invention is illustrated schematically in Fig 8. The arrangement, as mentioned earlier can be implemented as a computer unit or in a computer unit, comprising process units.

10   Thus, the arrangement comprises a processing unit 801 (such as a microprocessor of a computer), a module 802 for obtaining image masks, a sampling module 803, a calculating module 804, filtration module 805, classification machine 806 and a support vector machine and/or artificial neural network 807. As it is appreciated by a skilled person, one or several modules can be integrated together and/or in the

15   processor unit or run as instruction sets. Other units such as memories, interfaces etc. included for proper function of the computer unit are not illustrated.

It is appreciated that, the invention is not limited for signal processing of image data from generated in an x-ray apparatus. It is likewise possible to process any

20   image data seeking to find image information as described earlier.

It should be understood that the above-mentioned embodiment is only discussed for illustrative purposes and does not limit the invention. Numerous modifications and variations of the present invention are possible in light of the above teachings

25   without departing from the spirit and scope of the invention as limited only by the following claims.